

Does Gender Play a Role in the Assessment of Oral Proficiency?

Khalil Motallebzadeh

Department of English, Torbat-e-Heydareih Branch, Islamic Azad University

P.O.Box 140, Torbat-e-Heydareih, Iran

E-mail: k.motalleb@iautorbat.ac.ir; kmotallebz@gmail.com

Shaahin Nematizadeh (Corresponding author)

English Department, Islamic Azad University, Science and Research Branch, Iran

Flat 9, No. 17, Jaar St., Northern Mofatteh Ave. Tehran, Iran

Tel: 98-91-5510-4073 E-mail: Shaahin.nemati@gmail.com

Received: July 1, 2011

Accepted: July 19, 2011

Published: December 1, 2011

doi:10.5539/elt.v4n4p165

URL: <http://dx.doi.org/10.5539/elt.v4n4p165>

Abstract

Gender has been a controversial issue which affects the language learning process. McNamara (1996) has proposed that there are some variables affecting second language performance one of which is sex. In much the same way, it has been reported that gender plays a role in the area of language testing (Brown, 2003; Lumley & O'Sullivan, 2005; Motallebzadeh, 1993; O'Sullivan, 2002). The present study is, thus, an attempt to explore the possible relationship between gender and oral performance of Iranian intermediate and upper intermediate EFL language learners. For this purpose, 429 adult students in six different institutions in Mashhad and Kerman participated in the study. After the Oxford placement test and an IELTS-format oral placement test, 160 of them were selected for a final oral interview. Finally, through a T-test, it was found out that females did better in oral performance than males, however, the difference was not that significant.

Keywords: Gender, Oral Assessment, IELTS Speaking Test

1. Introduction

It has long been discussed that a second language learners' gender is likely to have some effects on the process of language learning and learner's performance in particular (Ellis, 1994; Brown, 2000). Whether or not such an effect is positive or negative is a frequently-debated subject for research. Knowing the possible effect of this variable might help language teachers and examiners avoid its interference in a reliable assessment. Dörnyei (2005) discusses that gender is such a variable which has been shown to play a significant role in the success of learners in the process of language learning and there is a considerable amount of literature on all dimensions of SLA affected by gender.

Knowing the possible effect of the gender of learners on the process of language learning and testing will certainly pave the way to better strategy and method selection in both language learning and teaching. Furthermore, factors influencing the performance of individuals in a test environment have been occasionally investigated. However, when it comes to the assessment of language oral ability, the point gets even more controversial, since assessing oral abilities and speaking in particular requires a completely different process. To support this fact, Fulcher (2003, cited in Brown, 2005) contributes to the debates about the validity which arises in relation to these tests. He also talks of "rater reliability and bias, how affective factors influence performance, the importance of washback, and the tension between linguistic competence and communicative ability" (p. 236).

2. Theoretical and Empirical Background

2.1 Concept of oral assessment

According to Underhill (1987), oral test is a "procedure in which the learner speaks and is assessed on the basis of what he says (p. 7)". Brown (2004) also discusses that interactive tasks which are subcategories of performance-based assessment "involve learners in actually performing the behavior that we want to measure. In interactive tasks, test takers are measured in the act of speaking, requesting, responding or in combining listening and speaking such as oral interviews" (p. 11).

2.2 Gender, language and assessment

Now focused attention will be paid to the genders' performance in some language contexts.

Research has come up with conclusions that male and females differ significantly in terms of their test-taking abilities (Brown, 2003; Lumley & O'Sullivan, 2005; Motallebzadeh, 1993; O'Sullivan, 2002). Chastain (1988) talks of an unpublished study comparing achievement scores of boys and girls in each of four language skills and found that girls' scores were higher in written skills while boys' scores were higher in oral skills. In the same regard, the UK assessment of performance unit (1986, cited in Cook, 2001) found English girls were better at French than boys in all skills except speaking. In addition, Coleman (1996, cited in Cook, 2001) states second language learning is more popular among girls and almost 70% percent of learners are females.

In support of this idea, Stumpf and Stanely (1998, cited in Lahey, 2001) discuss that women perform better than men in a range of language skills including "verbal and spatial memory, perceptual speed" (p. 11), whereas men perform better than women in "mathematics, science and social studies" (p. 11). Over the past decade, researchers such as Barton (2002, cited in Davies, 2004) have noted, in particular, that the disparity in performance between boys and girls is significantly greater in modern languages than in other areas of the curriculum" (p. 53).

Amazingly, researches have shown that there could be some differences in performances of males and females in language tests. Stumpf and Stanely (1998, cited in Lahey, 2003) state that, men generally receive higher scores on tests of spatial and mechanical reasoning. Lumley and O'Sullivan (2005) in a study to find whether there are effects on performance attributable to an interaction of variables such as the task topic, the gender of the person presenting the topic and the gender of the candidate, found that "the female students tended to slightly outperform male students, although the actual difference was not significant" (p. 434). O'loughlin (2002) who carried out a research on the effect of gender on oral proficiency testing, surprisingly did not find any significant difference in the performance of different genders. He also states that such researches have frequently met contradictory results and conjectures that the characteristics of contexts and the participants might simply be the source of this contradiction not necessarily the effect of gender in oral assessment.

Norton (2005), in her study to examine the advantages and disadvantages of paired-tasks for testing oral proficiency found out that, Japanese females paired with males of other nationalities would "adopt a floor-supporting role in the three-way discussion task by using more backchannelling tokens and allowing their male partners to take the floor first" (p. 294). This study also concluded that in 60% of the data samples where males were paired with females, male candidates produced more talk.

Markham (1988) carried out a study on gender bias in listening recall. Having males and females test-takers listen to introduced and un-introduced male and female speakers present a passage, he found out that female subjects recalled more idea units by listening to male speakers and that might be due to having been "conditioned to be more attentive to male speakers as a result of gender-related status divisions in the speech community" (p. 404).

A research conducted by O'Sullivan (2000) in which genders of interviewers were the focus of discussion revealed interesting results. O'Sullivan reports that:

Twelve Japanese learners were interviewed, once by a man and once by a woman. Video tapes of these interactions were scored by trained examiners. Comparison of scores indicated that in all scores except one case, the learners performed better when interviewed by a woman, regardless of the sex of the learner. (p. 373)

After analyzing the language produced by interviewers, systematic gender differences were found. Also it was concluded that as far as the interviewer is a female, the interviewees tended to produce more accurate language and when both interviewer and interviewee were females, the language produced was the most accurate among the other pairs.

On the contrary, Amjadian (2006) found that except for the pronunciation in which males did better, there were not significant differences between them. Also on written language tests such as Discrete point and Integrative tests, Motallebzadeh (1993) found that males would perform better in integrative tests than discrete point tests and that might be due to males' logical mind. He also found that males significantly outperformed females in reading comprehension and cloze tests (which are examples of integrative tests).

In a study specifically designed to investigate the effect of gender in native speaker/non-native speaker interaction, Gass and Varonis (1986, cited in Shehadeh, 1999) collected data from 20 NNS Japanese adults of English interacting on three communication tasks. "Men took greater advantage of the opportunities to use the conversation in a way that allowed them to produce a greater amount of comprehensible output, whereas women utilized the conversation to obtain a greater amount of comprehensible input" (Gass and Varonis, 1986, cited in Shehadeh, 1999,

p. 258).

Shehadeh (1999) carried out a research to explore gender differences in ESL classrooms. He gathered the required data from 16 males and 19 females adult subjects aging from 22 to 37. There were eight native speakers (four males and four females) and 27 non-native speakers of English (twelve males and fifteen females) most of whom were acquainted with each other as ESL classmates on the same course and represented 13 different first language (LI) backgrounds. The findings of Shehadeh's study support those reported by Gass and Varonis (1986) in that:

Men appeared to take greater advantage in the group activity (a mixed-sex task) to use the conversation in a way that allowed them to retain the turn, enjoy a greater amount of talk, and thus produce a greater amount of comprehensible output than women. But Shehadeh's study also revealed that same-sex dyads offered women comparatively greater opportunities to produce comprehensible output than men. It is not yet clear whether these differences in gender are innately/biologically determined, or psychologically and/or socio-culturally bound. (P. 257)

3. Method

3.1 Participants

429 language learners of intermediate levels at six different language institutes in Mashhad and Kerman, Iran served as the primary participants of this study. The participants aged from 15 to 49. For the purpose of homogenizing the participants, they were initially tested using the Oxford placement test. Choosing intermediate and upper-intermediate level participants (those who scored from 60% to 70% of the whole mark), the number of the participants was reduced to 198. The participants were for a second time assessed but orally by two instructed and experienced interviewers using IELTS speaking assessment descriptors. Selecting those who scored between 4 to 7 out of the 0-9 IELTS score bands, the number of participants was once again reduced to 160. The selection of the intermediate and upper-intermediate participants was based on the Common European Framework of Reference for languages (CEFR, 2001) scale. On this scale, intermediate and upper-intermediate levels are called B1 and B2. These 160 participants homogenized in terms of both linguistic knowledge and oral proficiency were used as the main participants of the study. The final stage of the study which was an oral interview lasting for about four to five minutes used materials and procedures of IELTS speaking part 2 and 3. The interviews were recorded to be rated later. The performances of the male and female participants were analysed, scored and compared to one another to see if genders differ from each other with regard to their oral proficiency. It should be mentioned that the whole process lasted for 32 days.

3.2 Design

In this research, the participants were different from each other on their gender which is an independent variable and the purpose was to examine if there exists a relationship between gender and oral proficiency (dependent variable). Thus, this study had an ex post facto design. Ex post facto design refers to a kind of research which tries to "find a relationship between the dependent and independent variables" (Hatch & Farhady, 1981, p. 26).

3.3 Instrumentation

The following instruments were utilized in this study to gather data on the participants' linguistic level and oral proficiency.

3.3.1 The Oxford placement test

The Oxford placement test filtered the participants and homogenized them in terms of proficiency level. It was administered in the institutions to pick up the participants of study. This test includes 50 items on the grammatical structures and the participants were allowed 25 minutes to complete it.

3.3.1.1 Reliability of the Oxford placement test

After the participation of 429 intermediate language learners, the reliability of the Oxford placement test was calculated using Cronbach's Alpha method. Through a pilot study using 40 cases of the participants who were randomly selected, the reliability turned out to be 0.787.

3.3.2 Oral placement interviews

The participants were first interviewed using IELTS speaking part one (questions based on personal information) to be selected as the research participants. The participants considered intermediate and upper-intermediate (scoring between 4 to 7 on the 1-9 IELTS scale) in terms of oral proficiency were selected based on the IELTS speaking assessment descriptors (Public Version).

3.3.3 IELTS format oral interviews

Participants finally underwent the main oral interviews using IELTS speaking parts 2 and 3 conducted by two different trained and instructed interviewers. The interviewees were first given an IELTS speaking prompt card and a minute to think and take notes on the cards' content. They were then asked to speak about the subject for 2 minutes (part 2). After approximately two minutes, the interviewer would start a related discussion on the same prompt with the interviewee (part 3).

3.3.4 IELTS speaking assessment descriptors (public version)

IELTS speaking assessment descriptors were taken into account by both interviewers and raters for assessing participants' oral abilities.

3.3.5 Raters training session

The raters were experienced IELTS teachers and had personally received the overall scores of 8 and 7.5, respectively on the official IELTS test. Despite these facts, two raters training sessions were held each lasting for 60 minutes to get the raters more familiar with the IELTS speaking assessment descriptors. The raters were provided with a copy of descriptors and asked to study the descriptors carefully prior to the first training session. To have a harmonious approach towards the descriptors, they were discussed, analysed, and clarified by the raters. Over 15 recordings of IELTS interviewes and 7 videos taken from available IELTS books in the market and downloaded from the internet were used as the training session materials. The raters would score each recording and reason why they assigned a particular score to each. This process lasted for two sessions so that the raters could come to a fairly logical and unanimous understanding of the assessment descriptors.

3.4 Procedure

The following procedure was carried out to conduct the research. First of all, the Oxford placement test was administered to 429 learners studying at intermediate courses in six different institutions in Mashhad and Kerman. As a result of this test, the number of participants was reduced to 198.

In order for the participants to be homogeneous in terms of their language oral abilities, they were each interviewed orally using the IELTS speaking part 1 format which lasted for about 5 minutes. This test which played the role of an oral placement test, reduced the number of participants to 160.

Immediately after the oral placement test, the participants were once again assessed orally using the IELTS speaking part 2 and 3 formats. Part 2 and 3 lasted approximately for 2 minutes each. The interviews were recorded to be listened to and rated at a later time. Afterwards, the recordings were assessed by two different raters. Raters assessed the recordings based on four IELTS assessment criteria. They include fluency and coherence, lexical resource, grammatical accuracy, and pronunciation. (O'Connell, 2006)

4. Results

At the beginning of this study, the fifty-item Oxford placement test was administered to 429 intermediate language learners. The purpose was to homogenize the participants in terms of their linguistic structural knowledge. Table 1 shows the descriptive statistics of the participants who agreed to take part in this study. As the table shows, the mean score is 27.53 out of 50. The highest score achieved was 44 and the lowest was 11. The most repeated score (mode) was 30.

Insert Table 1 Here

As it can be observed in Table 2, the participants who received less than 60% or more than 70 % of the whole mark were crossed out from the study. It should be mentioned that those participants who were 16 or younger and 33 or older were also omitted from the selected population in order to have adult as the target group age of the research. Selecting the adult intermediate and upper-intermediate level participants, the number of participants was reduced to 198.

Insert Table 2 Here

As it is observable in Figure 1, the participants were once again reduced to 160 as a result of an oral placement test. This test lasted for 4 to 5 minutes and used materials from IELTS speaking interview part 1. The interviewers rated the participants based on the IELTS speaking assessment descriptors. During the oral placement sessions, only the participants who scored 4 to 7 on the 1-9 scale were selected as the final research participants.

Insert Figure 1 Here

After administering the written and oral placement tests, the researcher came up with 160 participants homogenized in terms of written and oral proficiency. About 63% percent of the participants were females and 37% were males.

Table 3 illustrates the number and genders of the participants. (Insert Table 3 Here)

The final 160 participants went through IELTS part 2 and 3. The interviewees were given a prompt card and a minute to get prepared. They would then be asked to speak about the prompt and answer some questions related to the prompt. The interviews were recorded to be rated by more than one rater. It should be noted that raters were trained in two 90-minute training sessions. Table 4 illustrates the scores given by the first rater. The lowest score assigned is 3 and the highest is 7. The mean, mode, and standard deviation are 5.57, 5.5, and 0.64, respectively. Table 5 provides information about the assessment of the second rater. The scores in the second ratings range from 3 to 7.5. The mean is 5.03, the mode is 5, and the standard deviation is 0.7. Table 6 shows that the inter-rater reliability was calculated using Cronbach's Alpha method which turned out to be 0.74.

Insert Table 4, 5 and 6 Here

To determine if genders differ from one another in terms of their oral performance, a T-test was conducted. Table 5 shows the information related to the number of males and females and the means of their scores in the oral interviews. It should be mentioned that the underlying assumption of the independent-test which is homogeneity of the variances of the two groups, is met. (Insert Table 7 Here)

5. Discussions and Conclusions

In order to compare males and females in terms of their oral performance, the means of each group were compared through a T-test. By doing so, the null hypothesis of the study was rejected. Genders performed differently in oral interviews with females performing slightly better.

Throughout this study, after analysing the oral performance of male and female participants, it was found out that gender plays not a very significant role in the oral assessment process which is quite consistent with the study carried out by O'Loughlin (2002). This slight difference in performance of genders might have originated from the more serious look of females at the learning process. Measures need to be taken to get males more interested, motivated, and serious in language learning classes. Considering that paired oral tests might reduce this difference, in class pair-works and male/female interactions might reduce the difference to zero. It was also noticed that when the female participants noticed they were being recorded, they seemed to be stressed out which influenced their performance, while this was less observed with male participants. Male/female interactions in class would also help increase the females' confidence. In addition, one way to overcome the effect of stress on oral interviews might be paired oral tests. Paired oral tests have been found to reduce the stress and provide a relaxing environment for interviews. (Foot, 1999; Seville & Hargreaves, 1999, cited in Norton, 2005).

Although age was not a variable to be discussed in this study, it is worth mentioning that there was an opposite interaction between age and oral performance. The younger the participants were, the better performance was observed. In contrast, the scores gradually decreased as the age of participants increased.

References

- Amjadian, M. (2004). The relationship between the acquaintanceship with the interviewer, the interviewee's sex and Iranian EFL performance in oral interview as a test technique. *Unpublished master's thesis*. Tarbiat Modarres University, Tehran, Iran.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20 (1), 1-25. <http://dx.doi.org/10.1191/0265532203lt242oa>
- Brown, A. (2005). Book review: Testing second language speaking. *Language Testing*, 22(2), 235-241. <http://dx.doi.org/10.1177/026553220502200206>
- Brown, D. (2000). *Principles of language learning and teaching* (3rd ed.). New York: Pearson Education.
- Brown, D. (2004). *Language assessment, principles and classroom practices*. New York: Pearson Education.
- CEFR. [Online] Available: http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf (July 12, 2011)
- Chastain, K. (1988). *Developing second language skills*. Florida: Harcourt Brace Jovanovich.
- Cook, V. (2001). *Second language learning and language teaching* (3rd ed.). New York. Oxford University Press Inc.
- Davies, B. (2004). The gender gap in modern languages: a comparison of attitude and performance in year 7 and year 10. *Language Learning Journal*, 21, 53-58
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. New Jersey: Lawrence Erlbaum Associates.

- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Foot, M. C. (1999). Relaxing in pairs. *ELT Journal*, 53(1), 36 – 41. <http://dx.doi.org/10.1093/elt/53.1.36>
- Hatch, E., & Farhady, H. (1982). *Research design and statistics*. CA: Newbury House.
- Lahey, B. B. (2001). *Psychology: An introduction*. New York: McGraw-Hill Education.
- Lumley, T., & O’Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing*, 22(4), 415-437. <http://dx.doi.org/10.1191/0265532205lt303oa>
- Markham, P. (1988). Gender and perceived expertness of the speaker as factors in ESL listening recall. *Tesol Quarterly*, 22, 397- 405
- McNamara, T. (1996). *Measuring second language performance*. Harlow: Longman.
- Motallebzadeh, K. (1993). On the relationship between DP/IN tests of language proficiency and learners’ sex. *Unpublished master’s thesis*. Tarbiat Modarres University, Tehran, Iran.
- Norton, J. (2005). The paired format in Cambridge speaking tests. *ELT Journal*, 59 (4), 287-297
- O’Connell, S. (2006). *Focus on IELTS foundation*. Essex: Pearson Education.
- O’Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, 19(2), 169-92. <http://dx.doi.org/10.1191/0265532202lt226oa>
- O’Sullivan, B. (2002). Exploring gender and oral proficiency interview performance. *System*, 28, 373-386
- Shehadeh, A. (1999). Gender differences and equal opportunities in the ESL classroom. *ELT Journal Volume*, 53(4), 256-261
- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.

Table 1. Descriptive Statistics of 429 participants taking the Oxford placement test.

N	Valid	198
N	Missing	0
Mean		31.84
Median		31.50
Mode		30
Variance		3.131
Minimum		30
Maximum		35

Table 2. Descriptive statistics of 198 qualified after the Oxford placement test.

N	Valid	429
N	Missing	0
Mean		27.53
Median		30
Mode		30
Variance		35.773
Minimum		11
Maximum		44

Table 3. Number and gender of the participants after homogeneity tests.

Gender	Frequency	Valid Percent
Female	101	63.1
Male	59	36.9
Total	160	100.0

Table 4. Frequency of the scores assigned by the first rater.

Score Bands	Frequency	Valid Percent
3.00	1	.6
4.00	5	3.1
4.50	9	5.6
5.00	24	15.0
5.50	59	36.9
6.00	39	24.4
6.50	212	13.1
7.00	2	1.3
Total	160	100.0

Table 5. Frequency of the scores assigned by the second rater.

Score Bands	Frequency	Valid Percent
3.00	1	.6
3.50	2	1.3
4.00	16	10.0
4.50	40	25.0
5.00	49	30.6
5.50	30	18.8
6.00	12	7.5
6.50	4	2.5
7.00	4	2.5
7.50	2	1.3
Total	160	100.0

Table 6. Inter-rater reliability.

Cronbach's Alpha	Number of Items	N of cases
0.74	2	160

Table 7. Group statistics of oral scores and mean differences.

Sex	N.	Mean	Std. Deviation	Std Error Mean
Female	101	5.8069	.54300	.05403
Male	59	5.3263	.54388	.07081

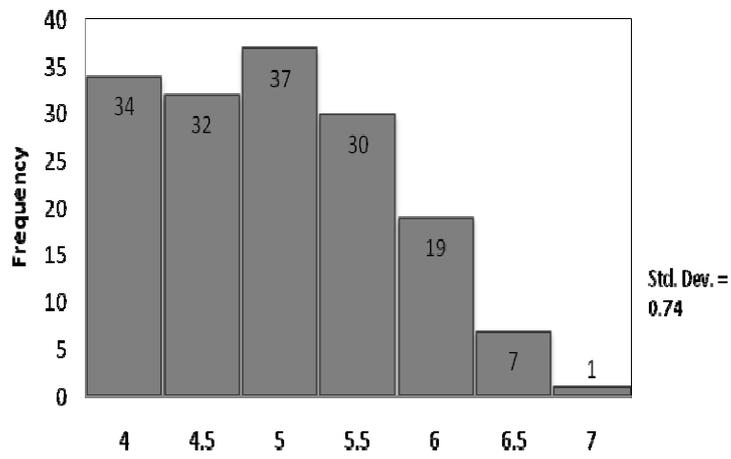


Figure 1. Oral placement scores frequency